REMARKS

In the Claims:

Claim 27 has been broadened to clarify that the wild type sequence may be found in any tissue, as claimed in originally filed claim 27. Original claim 27 has previously been examined in the Office Action dated March 24, 2003. Claims 31 and 33 have been amended to correct dependency. Specifically, claims 31 and 33 depended from canceled claim 22. As amended, claims 31 and 33 correctly depend from claim 25. Claim 35 has been canceled herein without prejudice or disclaimer.

Claim Rejections:

35 U.S.C. § 101

The Examiner has maintained his rejection of claims 25-34 and new claim 36 is rejected under 35 U.S.C. § 101. The Examiner alleges that the claimed invention is not supported by either a substantial asserted utility or a well established utility. Specifically, the Examiner does not find Applicants arguments, based on the Goddard and Ashkenazi declarations to be persuasive. Applicants respectfully disagree and maintain that the Goddard and Ashkenazi declarations demonstrate that the claimed invention is supported by both a specific, substantial, asserted utility and a well established utility.

As argued in the response filed December 24, 2003, the totality of the evidence demonstrates that the present claims are supported by a specific and substantial utility. Specifically, even if one assumes *arguendo* that it is more likely than not that there is no correlation between gene amplification and increased mRNA/protein expression, a polypeptide encoded by a gene that is amplified in cancer would still have a specific and substantial utility. As Dr. Ashkenazi explains at paragraph 6 of his previously filed declaration:

> [E]ven when amplification of a cancer marker gene does not result in significant over-expression of the corresponding gene product, this very absence of gene product over-expression of the corresponding gene product still provides significant information for cancer diagnosis and treatment. Thus, if over-

expression of the gene product does not parallel gene amplification in certain tumor types but does so in others, then parallel monitoring of gene amplification and gene product over-expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over-expression is crucial information for the practicing clinician. If a gene is amplified but the corresponding gene product is not over-expressed, the clinician accordingly will decide not to treat a patient with agents that target that gene product.

Accordingly, the PRO347 polypeptide has a substantial, specific utility.

In any event, "[a]n applicant need only provide one credible assertion of specific and substantial utility for each claimed invention to satisfy the utility requirement." MPEP § 2107. One of ordinary skill in the art would appreciate that the claimed polypeptides would be useful in creating degenerative oligonucleotide probes, which can be used to isolate genomic and cDNA nucleotide sequences. *See* Lodish *et. al., Molecular Cell Biology* (4th ed.) 2000, Figure 7-19 (Appendix A).

Specific oligonucleotide probes for the gene encoding PRO347 can by synthesized chemically based on a portion of the amino acid sequence disclosed in SEQ ID NO:50. For an oligonucleotide to be useful as a probe, it must be long enough for its sequence to occur uniquely in the clone of interest and not in any other clones. The skilled artisan would appreciate that generally that means the oligonucleotide should contain at least 20 nucleic acids. In principle, probes of that length may be prepared from a sequence as short as 7 amino acids.

Degenerative oligonucleotide probes may be created by digesting PRO347 with one or more proteases into specific peptides. Based on the genetic code, the oligonucleotide sequences encoding the determined peptide sequences can be predicted. Due to the degeneracy of the genetic code, oligonucleotides containing all possible combinations of codons are synthesized to assure that at least one will match the gene perfectly. The final step is to radiolabel the oligonucleotides by transferring a ^{32}P-labeled phosphate group from ATP to the 5' end of each oligonucleotide.

Screening a cDNA library with a degenerate probe (created by using the polypeptide of PRO347) using membrane-hybridization techniques, will identify PRO347 nucleic acid

6

clones that hybridize to the perfectly complementary oligonucleotide present in the probe mixture. Thus, use of the polypeptide sequence of PRO347 for creating oligonucleotide probes is a specific utility. "Specific utility" is defined as:

> [a] utility that is *specific* to the subject matter claimed. This contrasts with a *general* utility that would be applicable to the broad class of the invention. For example, a polynucleotide whose use is disclosed simply as a 'gene probe' or 'chromosome marker' would not be considered to be *specific* in the absence of a disclosure of a specific DNA target. Similarly, a general statement of diagnostic utility, such as diagnosing an unspecified disease, would ordinarily be insufficient absent a disclosure of what condition can be diagnosed.

Revised Interim Utility Guidelines Training Materials, pgs. 5-6 (http://www.uspto.gov/web/offices/pac/utility/utilityguide.pdf). Utility of the presently claimed polypeptide for generating oligonucleotide probes is specific because such an oligonucleotide probe would be useful for identifying and isolating the nucleic acid encoding the claimed polypeptide. It is undisputed that the nucleic acid (SEQ ID NO:49) encoding the claimed polypeptide is amplified in lung and colon tumors and therefore, has a specific diagnostic utility. *See* pages 119-137 of the specification describing the gene amplification experiments. Thus, the presently claimed polypeptide has the specific utility of generating a probe capable of identifying and isolating the nucleic acid encoding it, where the nucleic acid has a diagnostic utility because it is amplified in lung and colon tumors.

Use of the polypeptide sequence of PRO347 for creating oligonucleotide probes is also a substantial utility. "Substantial utility" is defined as:

> a utility that defines a 'real world' use. Utilities that require or constitute carrying out further research to identify or reasonably confirm a "real world" context of use are not substantial utilities. For example, both a therapeutic method of treating a known or newly discovered disease and an assay method for identifying compounds that themselves have a "substantial utility" define a "real world" context of use. An assay that measures the presence of a material which has a stated correlation to a predisposition to the onset of a particular disease condition would also define a "real world" context of use in identifying potential candidates for preventive measure or further monitoring.

7

Revised Interim Utility Guidelines Training Materials, pg. 6http://www.uspto.gov/web/ offices/pac/utility/utilityguide.pdf). Utility of the presently claimed polypeptide for generating oligonucleotide probes is substantial because such an oligonucleotide probe would be useful for identifying and isolating the nucleic acid encoding the claimed polypeptide. It is undisputed that the nucleic acid (SEQ ID NO:49) encoding the claimed polypeptide is amplified in lung and colon tumors and therefore, has a substantial diagnostic utility. *See* pages 119-137 of the specification describing the gene amplification experiments. Thus, the presently claimed polypeptide has the substantial utility of generating a probe capable of identifying and isolating the nucleic acid encoding it, where the nucleic acid has a diagnostic utility because it is amplified in lung and colon tumors.

Further, use of the polypeptide sequence of PRO347 for creating oligonucleotide probes is a credible utility:

> Where an applicant has specifically asserted that an invention has a particular utility, that assertion cannot simply be dismissed by Office personnel as being 'wrong'. Rather, Office personnel must determine if the assertion of utility is credible (*i.e.,* whether the assertion of utility is believable to a person of ordinary skill in the art based on the totality of evidence and reasoning provided). An assertion is credible unless (A) the logic underlying the assertion is seriously flawed, or (B the facts upon which the assertion is based are inconsistent with the logic underlying the assertion. Credibility as used in this context refers to the reliability of the statement based on the logic and facts that are offered by the applicant to support the assertion of utility. A *credible* utility is assessed from the standpoint of whether a person of ordinary skill in the art would accept that the recited or disclosed invention is currently available for such use. For example, no perpetual motion machines would be considered to be currently available. However, nucleic acids could be used as probes, chromosome markers, or forensic or diagnostic markers. Therefore the credibility of such an assertion would not be questioned, although such a use might fail the *specific* and *substantial* tests.

Revised Interim Utility Guidelines Training Materials, pg. 5 (http://www.uspto.gov/web/ offices/pac/utility/utilityguide.pdf). Utility of the presently claimed polypeptide for generating oligonucleotide probes is credible because neither the logic underlying this assertion is flawed, nor are the facts inconsistent with this logic. Moreover, the Utility Guidelines specifically recognize probes as having a credible utility.

Finally, use of the polypeptide sequence of PRO347 for creating oligonucleotide probes is a well-established utility. A "well established" utility is a:

> specific, substantial, and credible utility which is well known, immediately apparent, or implied by the specification's disclosure of the properties of the material, alone or taken with the knowledge of one skilled in the art.

Revised Interim Utility Guidelines Training Materials, pg. 7 (http://www.uspto.gov/web/offices/pac/utility/utilityguide.pdf). Utility of the presently claimed polypeptide for generating oligonucleotide probes is well established because it is implied by the specification's disclosure, and would be immediately apparent to one of ordinary skill in the art. At page 16, lines 3-5 of the specification, Applicants assert that "oligonucleotide probes (are) useful for isolating genomic and cDNA nucleotide sequences or as antisense probes." The gene amplification experiment, described at pages 119-137 of the application, demonstrates the specific diagnostic utility (for lung and colon tumors) of the nucleic acid encoding PRO347. At the time the application was filed, it was well known to one of ordinary skill in the art to use a polypeptide sequence to generate oligonucleotide probes capable of identifying and isolating a nucleic acid sequence of interest, such as the nucleic acid sequence encoding PRO347. *See* Lodish *et. al., Molecular Cell Biology* (4th ed.) 2000, Figure 7-19 and Chapter 7.3 (*Appendix A*).

For all the above reasons, Applicants have demonstrated currently pending claims 25-34, and 36 are supported by an asserted substantial, specific, and well-established utility. Therefore, Applicants respectfully request that the Examiner withdraw the rejection of claims 25-34, and 36 for lack of utility.

35 U.S.C. § 112 ¶ 1, Enablement-Utility

The Examiner has rejected claims 25-34, and 36 under 35 U.S.C. § 112 ¶1, alleging that because the claimed invention is not supported by either a specific asserted utility or a well established utility, one skilled in the art would not know how to use the claimed invention. As discussed in the remarks above, addressing the rejection under 35 U.S.C. § 101 for lack of utility, Applicants respectfully submit that the claimed polypeptide has the specific, substantial, and credible utility of generating degenerative oligonucleotide

9

probes, which can be used to isolated the amplified genomic and cDNA nucleotide sequences discussed in the gene amplification experiments at pages 119-137 of the specification. Further, Applicants disclose, at Example 11, pages 103, that protein sequences were used to search various databases and then used to create a consensus DNA sequence that was used as a probe to isolate a clone of the full-length coding sequence of PRO347.

Moreover, generation and use of degenerative oligonucleotide probes was well known by the time the application was filed. For example, Watson *et al.,* describe using a 32 amino acid sequence from the urate oxidase protein to design two degenerate oligonucleotide probes which were then used in a PCR reaction that resulted in a unique probe capable of hybridizing to a full length cDNA for urate oxidase. *See* Watson, *et al., Recombinant DNA* (2nd ed.) 1992, Figure 7-7 (*Appendix B*).

The Examiner also alleges that "[e]ven if there were a patentable use for the protein of SEQ ID NO:50, variants of 80-99% identity or conservative amino acid substitutions or amino acid additions, deletions, or substitutions, would not be enabled because the specification has not taught one of ordinary skill in the art how to use them or fragments thereof."

First, the currently pending claims are only directed to variants with at least 95% sequence identity. Second all claimed variants must also be encoded by a nucleic acid that is amplified in lung or colon tissue. Therefore, all claimed variants might also be used to create degenerative oligonucleotide probes capable of identifying and isolating nucleic acids that are overexpressed in lung or colon tumors.

Thus, one of ordinary skill in the art would be enabled to use the present invention to generate oligonucleotide probes capable of hybridizing to nucleic acid sequences that are amplified in lung and colon tumors. Applicants respectfully request the Examiner reconsider and withdraw the rejection of claims 25-34, and 36 under 35 U.S.C. § 112 ¶1 for their alleged inadequate disclosure on how to use the claimed invention.

35 U.S.C. § 112, first paragraph

Written Description

Claims 25-26, 33, and 34 remain rejected under 35 U.S.C. § 112, first paragraph as containing subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the art that the inventor(s), at the time the application was filed, had possession of the invention. In addition, the Examiner rejects claims 35 and 36 on this basis.

Applicants respectfully disagree. Claim 35 is cancelled herein without prejudice or disclaimer, and therefore, is not addressed here. Claims 25-26, 33, and 34 are adequately described in the present application. The Examiner contends that although the specification discloses a single polypeptide, there is no recited biological function for that protein, and therefore, it is not adequately described under 35 U.S.C. § 112, first paragraph. In reaching this conclusion, the Examiner relies on the analysis of the written description requirement set forth in the Guidelines for Examination of Patent Applications under 35 U.S.C. § 112, first paragraph (the "Guidelines").

However, the Guidelines support the conclusion that the written description requirement is satisfied for Claims 25-26, 33, and 34. Specifically, Claims 25 and 26 are adequately described under the analysis of Example 13, which sets forth an exemplary adequate written description for a claim to a genus of isolated proteins having a particular sequence, and which illustrates what is lacking in the description for a claim to genus of variants of the proteins having a particular sequence. The latter claim in Example 13 of the Training Materials is analogous to Claims 25, 26, and 36 of the present application. According to Example 13 of the Training Materials, a claim to "[a]n isolated variant of the protein of claim 1" is not adequately described if: (1) the specification and claim do not indicate what distinguishing attributes are shared by members of the claimed genus, (2) the specification and claim do not place any limit on the number of amino acid substitutions, deletions, insertions and/or additions, and (3) the specification and claim

11

fail to disclose structural features that could distinguish compounds in the genus from those outside the genus.

The Examiner's rejection of Claims 25, 26, and 36 is not proper because the present specification and claim *do* indicate distinguishing attributes that are shared by members of the claimed genus. For example, the specification and claim indicate the distinguishing attribute of being encoded by a nucleic acid that is amplified in lung or colon tumors, which is shared by all members of the claimed genus. In addition, the present specification and claims do limit the number of substitutions, deletions, insertions, and/or additions by requiring all sequences within the genus to be 95-99% identical to SEQ ID NO:50. Finally, the present specification at page 103 and Figure 20 discloses several structural features common to species falling within the claimed genus. Hence, Example 13 of the Guidelines indicates that the present disclosure adequately describes the claimed genus.

Further, the additional features of claims 33 and 34 are described at pages 62-63 of the specification. Therefore, Applicants respectfully request the Examiner withdraw this ground of rejection for claims 25-26, 33-34, and 36.

35 U.S.C. § 102:

Claims 25-34 remain rejected and new claims 35 and 36 are rejected under 35 U.S.C. § 102(a) as being anticipated by Botstein *et al.,* WO 99/35170, July 15, 1999 and Holtzman, WO 99/54343, October 28, 1999. Claims 25-27, 31, 33, and 34 remain rejected under 35 U.S.C. § 102(e) in light of Holtzman U.S. Patent Application Publication US20020028508. The Examiner maintains rejection of these claims under 35 U.S.C. § 102 on the basis of the Examiner's contention that Applicants have not overcome the utility rejection.

Applicants respectfully submit that they have overcome the utility rejection based on arguments presented in the present response. Therefore, Applicants' present

response, in addition to the response mailed 12/24/03, overcome the Examiner's rejection under 35 U.S.C. § 102. Applicants respectfully request that this rejection be withdrawn.

35 U.S.C. § 112, Second Paragraph

Claims 33-35 are rejected under 35 U.S.C. § 112, second paragraph as being indefinite for failing to particularly point out and distinctly claim the subject matter which Applicant regards as the invention. Specifically, Claims 33-34 are indefinite because they depend upon a canceled claim. Herein, Applicants have amended the dependency of claims 31 and 33 such that they no longer depend from canceled claim 22, but rather depend from claim 25. Claim 34 properly depends from claim 33.
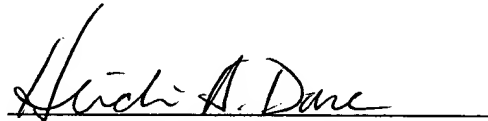
Claim 35 is cancelled herein without prejudice or disclaimer. Hence, Applicants have overcome this ground of rejection from claims 33-35 and respectfully request the rejection be withdrawn.

SUMMARY

Applicants believe that currently pending Claims 25-34, and 36 are patentable. Applicants respectfully request the Examiner grant allowance of this application. The Examiner is invited to contact the undersigned attorneys for the Applicant via telephone if such communication would expedite this application.

Respectfully submitted,

Heidi A. Dare
Registration No. 50,775
Attorney for Applicant

BRINKS HOFER GILSON & LIONE
P.O. BOX 10395
CHICAGO, ILLINOIS 60610
(312) 321-4200

APPENDIX A

MOLECULAR CELL BIOLOGY

Lodish Berk Zipursky Matsudaira Baltimore Darnell

W H FREEMAN AND COMPANY

Short Contents | Full Contents

Other books @ NCBI

Molecular Cell Biology

Harvey Lodish
Arnold Berk
Lawrence S. Zipursky
Paul Matsudaira
David Baltimore
James Darnell

Media Connected ⬆TOP

Molecular Cell Biology

Search

[GO]

(•) This book   ( ) All books

( ) PubMed

MOLECULAR CELL BIOLOGY

Lodish Berk Zipursky Matsudaira Baltimore Darnell

W H FREEMAN AND COMPANY

Short Contents | Full Contents

Other books @ NCBI

Molecular Cell Biology → **7. Recombinant DNA and Genomics** → 7.3. Identifying, Analyzing, and Sequencing Cloned DNA
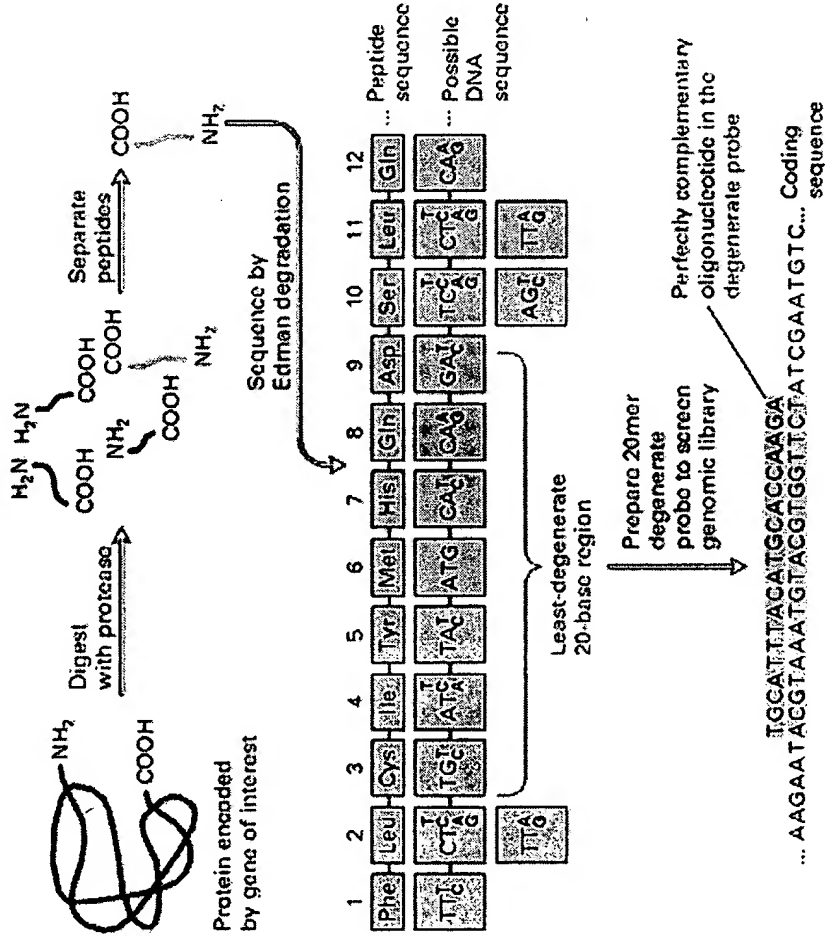
Figure 7-19. Designing oligonucleotide probes based on protein sequence. An isolated protein is digested with a selective protease such as trypsin, which specifically cleaves peptide bonds on the

Designing oligonucleotide probes based on protein sequence.

carboxy-terminal side of lysine and arginine residues. The resulting peptides are separated, and several are partially sequenced from their N-terminus by sequential Edman degradation. The determined sequences then are analyzed to identify the 6- or 7-aa region that can be encoded by the smallest number of possible DNA sequences. Because of the degeneracy of the genetic code, the 12-aa sequence (light green) shown here theoretically could be encoded by any of the DNA triplets below it, with the possible alternative bases at the same position indicated. For example, Phe-1 is encoded by TTT or TTC; Leu-2 is encoded by one of six possible triplets (CTT, CTC, CTA, CTG, TTA, or TTG). The region with the least degeneracy for a sequence of 20 bases (20-mer) is indicated by the red bracket. There are 48 possible DNA sequences in this 20-base region that could encode the peptide sequence 3 – 9. Since the actual sequence of the gene is unknown, a degenerate 20-mer probe consisting of a mixture of all the possible 20-base oligonucleotides is prepared. If a cDNA or genomic library is screened with this degenerate probe, the one oligonucleotide that is perfectly complementary to the actual coding sequence (blue) will hybridize to it.
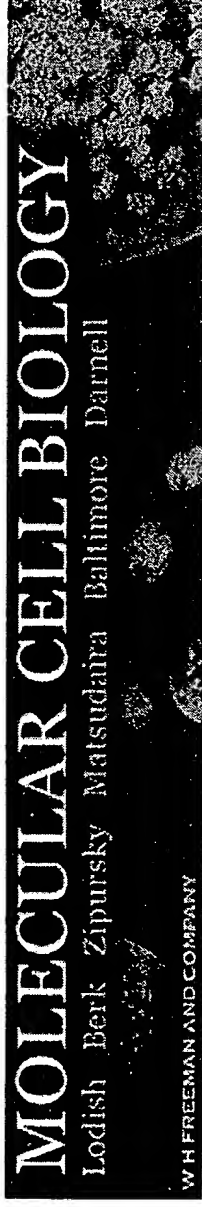
PERSPECTIVES for the Future

PERSPECTIVES in the Literature

Testing Yourself on the Concepts

MCAT/GRE-Style Questions

References

Search

Go

⦿ This book ◯ All books
◯ PubMed

© 2000 by W. H. Freeman and Company.

MOLECULAR CELL BIOLOGY

Lodish Berk Zipursky Matsudaira Baltimore Darnell

W H FREEMAN AND COMPANY

Other books @ NCBI

7.3. Identifying, Analyzing, and Sequencing Cloned DNA

Suppose you have isolated a particular protein and want to isolate the gene that encodes it. A complete genomic λ library from mammals contains at least a million different clones; a cDNA library must contain as many clones to include the sequences of scarce mRNAs. How are specific clones of interest identified in such large collections? The most common method involves screening a library by hybridization with radioactively labeled DNA or RNA probes. In an alternative method, a specific clone in a library of cloned DNA is identified based on some property of its encoded protein.

Once a particular genomic DNA or cDNA clone of interest has been identified and isolated from other clones, the cloned DNA can be separated from the vector DNA and analyzed. This separation is achieved by cleaving the recombinant vector with the same restriction enzyme used to insert the DNA fragment initially. During ligation of a cut vector and DNA fragments generated with the same restriction enzyme, the restriction recognition sequence is regenerated between the DNA fragments and vector (see Figure 7-8). Subsequent treatment with the same restriction enzyme will cut the recombinant vector at the same sites, releasing the vector and cloned DNA, which then can be separated by gel electrophoresis.

The most complete characterization of a cloned DNA requires determination of its nucleotide sequence; from this sequence, the amino acid sequence of an encoded protein can be deduced. The sequence of genomic DNA includes introns as well as exons; it also includes regions that control gene expression by determining the type of cell in which the encoded protein is expressed, the stage in development it is expressed, and the amount of protein produced. Genomic DNAs also include replication origins and sequences important in determining how the DNA associates with proteins in chromosomes. In subsequent chapters, we consider how cells use DNA sequences for these functions. In this section, techniques for identifying, characterizing, and finally sequencing cloned DNA are outlined.

Libraries Can Be Screened with Membrane-Hybridization Assay

As discussed in Chapter 4, under the conditions of temperature and ion concentration found in cells, DNA is maintained as a duplex (double-stranded) structure by the hydrogen bonds between A · T and G · C base pairs (see Figure 4-4b). DNA duplexes can be denatured (melted) into single strands by heating them in a dilute salt solution (e.g., 0.01 M NaCl), or by raising the pH above 11. If the temperature is lowered and the ion concentration in the solution is raised, or if the pH is lowered to neutrality, the A · T and G · C base pairs re-form between complementary single strands (see Figure 4-8). This process goes by many names: renaturation, reassociation, hybridization, annealing. In a mixture of nucleic acids, only complementary single strands (or strands containing complementary regions) will reassociate; the extent of their reassociation is virtually unaffected by the presence of noncomplementary strands. Such *molecular hybridization* can take place between two complementary strands of either DNA or RNA, or between an RNA strand and a DNA strand.

To detect specific DNA clones by molecular hybridization, cloned recombinant DNA molecules are denatured and the single strands attached to a solid support, commonly a nitrocellulose filter or treated nylon membrane (Figure 7-17). When a solution containing single-stranded nucleic acids is dried on such a membrane, the single strands become irreversibly bound to the solid support in a manner that leaves most of the bases available for hybridizing to a complementary strand. Although the chemistry of this irreversible binding is not well understood, the procedure is very useful. The membrane is then incubated in a solution containing a radioactively labeled single-stranded DNA (or RNA) *probe* that is complementary to some of the nucleic acid bound to the membrane. Under hybridization conditions (near neutral pH, 40 – 65 °C, 0.3 – 0.6 M NaCl), this labeled probe hybridizes to the complementary nucleic acid bound to the membrane. Any excess probe that does not hybridize is washed away, and the labeled hybrids are detected by autoradiog-raphy of the filter.

The procedure for screening a λ library with this membrane-hybridization technique is outlined in Figure 7-18. The recombinant λ virions present in plaques on a lawn of *E. coli* are transferred to a nylon membrane by placing the membrane on the surface of the petri dish. Many of the viral particles in each plaque adsorb to the surface of the membrane, but many virions remain in the plaques on the surface of the nutrient agar in the petri dish. In this way a *replica* of the petri dish containing a large number of individual λ clones is reproduced on the surface of the membrane. The original petri dish is refrigerated to store the collection of λ clones. The membrane is then incubated in an alkaline solution, which disrupts the virions, releasing and denaturing the encapsulated DNA. The membrane is then dried, fixing the recombinant λ DNA to the membrane's surface. Next, the membrane is incubated with a radiolabeled probe under hybridization conditions. Unhybridized probe is washed

away, and the filter is subjected to autoradiography.

The appearance of a spot on the autoradiogram indicates the presence of a recombinant λ clone containing DNA complementary to the probe. The position of the spot on the autoradiogram corresponds to the position on the original petri dish where that particular clone formed a plaque. Since the original petri dish still contains many infectious virions in each plaque, viral particles from the identified clone can be recovered for replating by aligning the autoradiogram and the petri dish and removing viral particles from the clone corresponding to the spot. A similar technique can be applied for screening a plasmid library in *E. coli* cells. ↑ TOP

Oligonucleotide Probes Are Designed Based on Partial Protein Sequences

Identification of specific clones by the membrane-hybridization technique depends on the ability to prepare specific, radiolabeled probes. Specific oligonucleotide probes for the gene encoding a protein of interest can be synthesized chemically if a portion of the amino acid sequence of the protein is determined. For an oligonucleotide to be useful as a probe, it must be long enough for its sequence to occur uniquely in the clone of interest and not in any other clones. For most purposes, this condition is satisfied by oligonucleotides containing about 20 nucleotides. This is because a specific 20-nucleotide sequence occurs once in every 4^{20} ($\approx 10^{12}$) nucleotides), a specific 20-nucleotide sequence in a genome usually occurs only one time. In principle, probes this length can be prepared based on only a 7-aa sequence out of a protein's total sequence (20 nucleotides ÷ 3 nucleotides per codon $\approx$ 7 amino acids). However, a somewhat longer amino acid sequence usually is determined to allow the preparation of several probes for use in cloning a gene of interest.

Here we describe two different approaches for preparing oligonucleotide probes. Generally, a radiolabeled oligonucleotide probe is used to screen a λ cDNA library using the membrane-hybridization technique. Once a cDNA clone encoding a particular protein is obtained, the full-length radiolabeled cDNA can be used to probe a genomic library for clones that contain fragments of the gene encoding the protein.

Degenerate Probes

One method for preparing a specific probe is outlined in Figure 7-19. The purified protein of interest is digested with one or more proteases (e.g., trypsin) into specific peptides, and the N-terminal amino acid sequences of a few of these peptides is determined by sequential Edman degradation or mass

Search

[Go]

○ This book   ○ All books
○ PubMed

Identifying, Analyzing, and Sequencing Cloned DNA

spectrometry (see Figures 3-46 and 3-47). Based on the genetic code, the oligonucleotide sequences encoding the determined peptide sequences can be predicted. Recall, however, that the genetic code is degenerate; that is, many amino acids are encoded by multiple codons (see Table 4-2). Since the specific codons used to encode the protein of interest are unknown, oligonucleotides containing all possible combinations of codons must be synthesized to assure that one of them will match the gene perfectly.

Once several peptides have been sequenced, the 6- or 7-aa stretch that can be encoded by the smallest number of possible DNA sequences is determined. For example, as illustrated in Figure 7-19, the amino acids in the sequence extending from position 3 through 8 (Cys-Ile-Tyr-Met-His-Gln) can be encoded by 2, 3, 2, 1, 2, and 2 possible codons, respectively. Consequently, 48 (= $2 \times 2 \times 1 \times 2 \times 2$) different 18-base DNA sequences could encode this one sequence of amino acids. The GA added at the 3′ end of these 18-base sequences must be complementary to the gene since the next amino acid in this peptide, Asp-9, is encoded by two codons that both start with GA. To be certain of obtaining a probe based on this amino acid sequence that hybridizes perfectly to the unique sequence present in the gene, all 48 of the 20-mer probes must be synthesized. A mixture of 20-mer probes based on any other portion of this peptide sequence would have to contain considerably more than 48 oligonucleotides because of the presence of leucine or serine residues, each encoded by six different codons.

A mixture of all the oligonucleotides that can encode a selected portion of a peptide sequence is called a *degenerate probe*. Such a mixture can be prepared at one time by adding more than one nucleotide precursor to the synthesis reaction at those points in the sequence that can be encoded by alternative bases. The final step in preparing this type of probe is to radiolabel the oligonucleotides, usually by transferring a ^{32}P-labeled phosphate group from ATP to the 5′ end of each oligonucleotide using *polynucleotide kinase* (Figure 7-20). Screening of a λ cDNA library with a degenerate probe using the membrane-hybridization technique will identify clones that hybridize to the perfectly complementary oligonucleotide present in the probe mixture. Under the usual experimental conditions, oligonucleotides that differ from the cDNA sequence at one or two bases also will hybridize. ↟ TOP

Unique EST-Based Probes

In recent years another approach has become available for obtaining a probe based on the partial amino acid sequence determined from an isolated protein. Because this approach utilizes cDNA sequence data, it identifies a single oligonucleotide, rather than a degenerate mixture, that can be used to screen a library for a particular gene. Using methods for DNA sequencing described later,

researchers have sequenced portions of vast numbers of cDNAs isolated from human cells and some additional important model organisms such as the mouse, *Drosophila*, and the roundworm *Caenorhabditis elegans*. These partial cDNA sequences, generally 200 to 400 bp in length, have been stored in computers and are available to researchers throughout the world via the Internet, the international computer network. This collection of partial cDNA sequences is called the *expressed sequence tag (EST)* database because it is composed of relatively short portions (tags) of genomic DNA sequence that are expressed in the form of mRNA. The EST database is constantly updated as sequences from increasing numbers of cDNA clones are added to it.

Computer programs that apply the genetic code are used to translate the EST sequences into partial protein amino acid sequences. Using programs that have been developed for the purpose and a personal computer, a researcher can search the current EST database for an EST that encodes a specific partial amino acid sequence in the particular protein under study. If a match is found, then the EST provides the unique DNA sequence of that portion of the full-length cDNA. A single, specific probe up to ≈100 bases long that is perfectly complementary to a portion of the EST can then be synthesized and radiolabeled. Alternatively, the polymerase chain reaction, described at the end of the chapter, can be used to synthesize a probe equal to the full length of the EST. By now, the human EST database is so large that ESTs can be identified that encode partial amino acid sequences determined from most isolated human proteins. ⬆ TOP

Specific Clones Can Be Identified Based on Properties of the Encoded Proteins

Genomic and cDNA libraries can also be screened for the properties of a specific protein encoded in the cloned DNA. This approach uses special cloning vectors, called λ expression vectors, in which the cloned DNA is transcribed into mRNA, which in turn is translated into the encoded protein. For example, λ phage vectors have been constructed so that the junction of inserted DNA lies in a region of the vector that is transcribed and translated at a high rate. Cloned DNA inserted at this position is transcribed into mRNA in every cell infected by this type of vector. If the cloned DNA contains a protein-coding sequence inserted in the same reading frame as the vector protein, infected cells will produce a *fusion protein* in which the amino terminus is encoded by the vector DNA and the remainder of the molecule by the cloned DNA (Figure 7-21).

When replica nitrocellulose filters are prepared from a recombinant library constructed in a λ expression vector, fusion proteins expressed from each individual clone are bound to the nitrocellulose filter. The replica filter can be screened by procedures capable of detecting specific fusion proteins. For example, a monoclonal antibody specific for a protein of interest can be incubated with replica

Identifying, Analyzing, and Sequencing Cloned DNA

filters of a λ cDNA expression library. If one of the λ clones expresses a fusion protein that includes the region of the protein bound by that monoclonal antibody, antibody molecules will bind to the filter at the position of that specific clone. After washing of the filter to remove unbound antibody, the position of the specific clone is detected by incubation with a second radioactively labeled antibody that recognizes the first antibody, followed by autoradiography of the filter.

In this method, termed expression cloning, any molecule that binds to a protein of interest with high affinity and specificity can be labeled and used as a probe to identify clones expressing the interacting protein. For instance, expression cloning has been useful in identifying cDNA clones encoding proteins that bind to specific DNA sequences; many such proteins are involved in controlling transcription. In this case, a labeled synthetic double-stranded DNA probe is incubated with replica filters prepared from a cDNA library cloned into a λ expression vector. Binding of the labeled DNA by fusion proteins locates the positions of desired clones on the original filter. As described in a later section, other types of expression vectors can be used to produce large amounts of a protein from a cloned gene. ↑TOP

Gel Electrophoresis Resolves DNA Fragments of Different Size

Once a specific DNA clone has been isolated, the cloned DNA is separated from the vector DNA by cleavage with the restriction enzyme used to form the recombinant plasmid, as described earlier. The cloned DNA and vector DNA then are separated by gel electrophoresis, a powerful method for separating proteins according to size (Chapter 3). Gel electrophoresis also is used to separate DNA and RNA molecules by size and to estimate the size of nucleic acid molecules of unknown length by comparison with the migration of molecules of known length.

DNA and RNA molecules are highly charged near neutral pH because the phosphate group in each nucleotide contributes one negative charge. As a result, DNA and RNA molecules move toward the positive electrode during gel electrophoresis. Smaller molecules move through the gel matrix more readily than larger molecules, so that molecules of different length, such as restriction fragments, separate (Figure 7-22). Because the gel matrix restricts random diffusion of the molecules, molecules of different length separate into "bands" whose width equals that of the well into which the original DNA mixture was placed. The resolving power of gel electrophoresis is so great that single-stranded DNA molecules up to about 500 nucleotides long can be separated if they differ in length by only 1 nucleotide. This high resolution is critical to the DNA-sequencing procedures described later. DNA molecules composed of up to ≈2000 nucleotides usually are separated electrophoretically on *polyacrylamide gels*, and molecules from 500 nucleotides to 20 kb on *agarose gels*.

Identifying, Analyzing, and Sequencing Cloned DNA

Two methods are common for *visualizing* separated DNA bands on a gel. If the DNA is not radiolabeled, the gel is incubated in a solution containing the fluorescent dye ethidium:
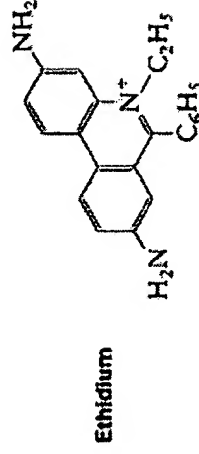
Ethidium

This planar molecule binds to DNA by intercalating between the base pairs. Binding concentrates ethidium in the DNA and also increases its intrinsic fluorescence. As a result, when the gel is illuminated with ultraviolet light, the regions of the gel containing DNA fluoresce much more brightly than the regions of the gel without DNA (Figure 7-23a).

Radioactively labeled DNA can be *visualized* by autoradiography of the gel. In this case, the gel is laid against a sheet of photographic film in the dark, exposing the film at the positions where labeled DNA is present. When the film is developed, a photographic image of the DNA is observed (Figure 7-23b). Radiolabeled DNA bands also can be detected by laying the gel against a phosphorimager screen, which counts β particles released by labeled molecules in the gel. The resulting data is stored by a computer and can be converted into an image of the gel that looks much like an autoradiogram. ⬆ TOP

Multiple Restriction Sites Can Be Mapped on a Cloned DNA Fragment

In addition to separating restriction fragments of different lengths, gel electrophoresis provides a means for estimating the length of fragments. As we've seen, the distance that a restriction fragment migrates in a gel is inversely proportional to the logarithm of its length. Thus the length of a restriction fragment can be determined fairly accurately by comparison with restriction fragments of known length subjected to electrophoresis on the same gel (see Figure 7-23a).

The ability to determine the length of restriction fragments makes it possible to locate the positions of restriction sites relative to one another on a DNA molecule (e.g., a newly cloned DNA fragment). A diagram showing the positions of restriction sites on a DNA molecule is called a *restriction map*, and the process of determining these positions is called *restriction-site mapping*. Figure 7-24 illustrates the procedure for mapping two restriction sites relative to each other when only *one* copy of each site is

Identifying, Analyzing, and Sequencing Cloned DNA

present in a fragment. In this simple case, three fragment samples are digested: one with enzyme I, one with enzyme II, and one with both enzymes.

When a DNA fragment contains *multiple* copies of the recognition site for one or more restriction enzymes, the mapping procedure is more complicated. In this case, the sites for each enzyme must be mapped before the sites for different enzymes can be mapped relative to one another. The first step in this procedure is radiolabeling the 5′ ends of both strands of the fragment with [γ-^{32}P]ATP and polynucleotide kinase (see Figure 7-20). As shown in Figure 7-25, the doubly end-labeled fragment is treated with a restriction enzyme that cuts the fragment just once, and the resulting singly end-labeled fragments are separated. These fragments then are partially digested with the enzyme whose multiple recognition sites are being mapped.

Multiple restriction sites in a cloned DNA fragment can be mapped by use of these two methods with multiple restriction enzymes. Because each distinct DNA sequence has a characteristic restriction-site map, such maps can be used to align partially overlapping cloned DNA fragments (see Figure 7-13). Also, specific small regions within a large cloned DNA fragment can be prepared by digesting the cloned fragment with various combinations of restriction enzymes; the smaller subfragments then can be isolated by gel electrophoresis. ⬆ TOP

Pulsed-Field Gel Electrophoresis Separates Large DNA Molecules

The gel electrophoretic techniques described so far can resolve DNA fragments up to ≈20 kb in length. Larger DNAs, ranging from 2×10^4 to 10^7 base pairs [20 kb to 10 megabases (Mb)] in length, can be separated by size with *pulsed-field gel electrophoresis*. This technique depends on the unique behavior of large DNAs in an electric field that is turned on and off (pulsed) at short intervals.

When an electric field is applied to large DNA molecules in a gel, the molecules migrate in the direction of the field and also stretch out lengthwise. If the current then is stopped, the molecules begin to "relax" into random coils. The time required for relaxation is directly proportional to the length of a molecule. The electric field then is reapplied at 90° or 180° to the first direction. Longer molecules relax less than shorter ones during the time the current is turned off. Since the molecules must relax into a random coil before moving off in a new direction, longer molecules start moving in the direction imposed by the new field more slowly than shorter ones. Repeated alternation of the field direction gradually forces large DNA molecules of different size farther and farther apart.

Pulsed-field gel electrophoresis is very important for purifying long DNA molecules up to ≈10^7 base

pairs in length (Figure 7-26). The technique is required for analyzing cellular chromosomes, which range in size from about 5×10^5 base pairs (smallest yeast chromosomes) to $2 - 3 \times 10^8$ base pairs (animal and plant chromosomes). Very large chromosomes must be digested into fragments of 10^7 base pairs or less before they can be analyzed. Such large restriction fragments can be generated with restriction enzymes that cut at rarely occurring 8-bp restriction sites. ↑ ᴛᴏᴘ

Purified DNA Molecules Can Be Sequenced Rapidly by Two Methods

Virtually all the information required for the growth and development of an organism is encoded in the DNA of its genome. The availability of techniques to produce and separate DNA restriction fragments a few hundred nucleotides long led to development of two procedures for determining the exact nucleotide sequence of stretches of DNA up to ≈500 nucleotides long. These DNA sequencing methods, together with the technology for constructing a library representing the entire genome of an organism, make it possible to determine the exact sequence of the entire DNA of that organism.

The total genomes of many viruses, several bacteria and archaeans, the yeast *S. cerevisiae*, and the roundworm *C. elegans* have already been sequenced. Automation of the techniques for sequencing DNA and computerized storage of the sequence data are facilitating the current effort to determine the sequence of the entire human genome. Within the next decade, if not sooner, researchers also are likely to complete sequencing the entire genomes of the fruit fly *Drosophila melanogaster*, the mouse, and other important experimental organisms. Knowledge of these DNA sequences will undoubtedly revolutionize our understanding of how cells and organisms function.

Maxam-Gilbert Method

In the late 1970s, A. M. Maxam and W. Gilbert devised the first method for sequencing DNA fragments containing up to ≈500 nucleotides. In this method, four samples of an end-labeled DNA restriction fragment are chemically cleaved at different specific nucleotides. The resulting subfragments are separated by gel electrophoresis, and the labeled fragments are detected by autoradiography. As illustrated in Figure 7-27, the sequence of the original end-labeled restriction fragment can be determined directly from parallel electrophoretograms of the four samples. ↑ ᴛᴏᴘ

Sanger (Dideoxy) Method

A few years later, F. Sanger and his colleagues developed a second method of DNA sequencing, which now is used much more frequently than the Maxam-Gilbert method. The Sanger method is also

Identifying, Analyzing, and Sequencing Cloned DNA

called *dideoxy sequencing* because it involves use of 2′,3′-dideoxynucleoside triphosphates (ddNTPs), which lack a 3′-hydroxyl group (Figure 7-28). In this method, the single-stranded DNA to be sequenced serves as the template strand for in vitro DNA synthesis; a synthetic 5′-end-labeled oligodeoxynucleotide is used as the primer.

As shown in Figure 7-29, four separate polymerization reactions are performed, each with a low concentration of one of the four ddNTPs in addition to higher concentrations of the normal deoxynucleoside triphosphates (dNTPs). In each reaction, the ddNTP is randomly incorporated at the positions of the corresponding dNTP; such addition of a ddNTP terminates polymerization because the absence of a 3′ hydroxyl prevents addition of the next nucleotide. The mixture of terminated fragments from each of the four reactions is subjected to gel electrophoresis in parallel; the separated fragments then are detected by autoradiography. The sequence of the original DNA template strand can be read directly from the resulting autoradiogram (see Figure 7-29c). Once the sequence for a particular cloned DNA fragment is determined, primers for overlapping fragments can be chemically synthesized based on that sequence. The sequence of a long continuous stretch of DNA thus can be determined by individually sequencing the overlapping cloned DNA fragments that compose it. ♠ ᴛᴏᴘ

SUMMARY

- • Generally, to isolate a DNA clone encoding a protein of interest, it is uniquely identified in a collection of a large number of different clones, such as a total genomic library or a cDNA library. Two common identification methods are (1) hybridization to a radiolabeled DNA probe specific for the clone and detection by autoradiography and (2) expression of the encoded protein and detection of the expressed protein by its biochemical activity or by its binding to a radiolabeled antibody specific for the protein.

- • A specific hybridization probe for cloned DNA encoding a protein of interest can be prepared based on a short portion of the amino acid sequence of the protein. Use of a degenerate probe, which is a mixture of all the possible DNA sequences that can encode the determined amino acid sequence, ensures that the radiolabeled probe contains the one sequence exactly complementary to the gene of interest (see Figure 7-19). When this perfectly complementary radiolabeled DNA probe hybridizes to DNA from the specific clone of interest on a replica filter, it can be detected by autoradiography (see Figure 7-18).

- • Vast numbers of partial cDNA sequences, or expressed sequence tags (ESTs), from humans and various model organisms are stored in computer databases. If a partial amino acid sequence is determined from an isolated protein, the EST database can be searched for an EST (i.e., a partial cDNA sequence) that encodes it. If found, the EST sequence can be used to prepare a perfectly complementary probe for isolation of the corresponding full-length cDNA from a

Identifying, Analyzing, and Sequencing Cloned DNA

cDNA library.

- • λ expression vectors are designed to express the protein encoded in cloned DNA fragments in *E. coli* cells infected with the vector. Plaques producing the encoded polypeptide can be detected with labeled molecules (e.g., an antibody) that bind to the protein of interest with high affinity (see Figure 7-21).

- • Gel electrophoresis separates DNA (and RNA) molecules according to their size. Consequently, a cloned DNA fragment, released from its cloning vector by digestion with the appropriate restriction enzyme, can be separated from the vector DNA by gel electrophoresis.

- • The mobility of either a double- or single-stranded DNA molecule during gel electrophoresis is inversely proportional to the logarithm of its length in nucleotides. Thus the size of a DNA molecule of unknown length can be determined by comparison to the electrophoretic migration of molecules of known length.

- • DNA molecules from 1 to 2000 nucleotides long are usually separated by electrophoresis in polyacrylamide gels; molecules from 500 nucleotides to 20 kb, by electrophoresis in agarose gels; and molecules from 20 to 10,000 kb, in pulsed-field agarose gels.

- • Multiple restriction sites for a particular restriction enzyme can be mapped in cloned DNAs by partially digesting end-labeled DNA molecules and then determining the length of the resulting fragments by gel electrophoresis and autoradiography (see Figure 7-25).

- • Once the cleavage sites for two restriction enzymes have been mapped on a DNA, these sites can be mapped relative to each other by comparing the sizes of DNA fragments generated with each enzyme separately and with both enzymes together (see Figure 7-24).

- • The sequence of a single-stranded DNA molecule can be determined by either the Maxam-Gilbert method or the Sanger dideoxy method. Sequences of up to ≈500 nucleotides can be determined on a single gel because DNA molecules up to this length can be separated on polyacrylamide gels when they differ in length by a single nucleotide.
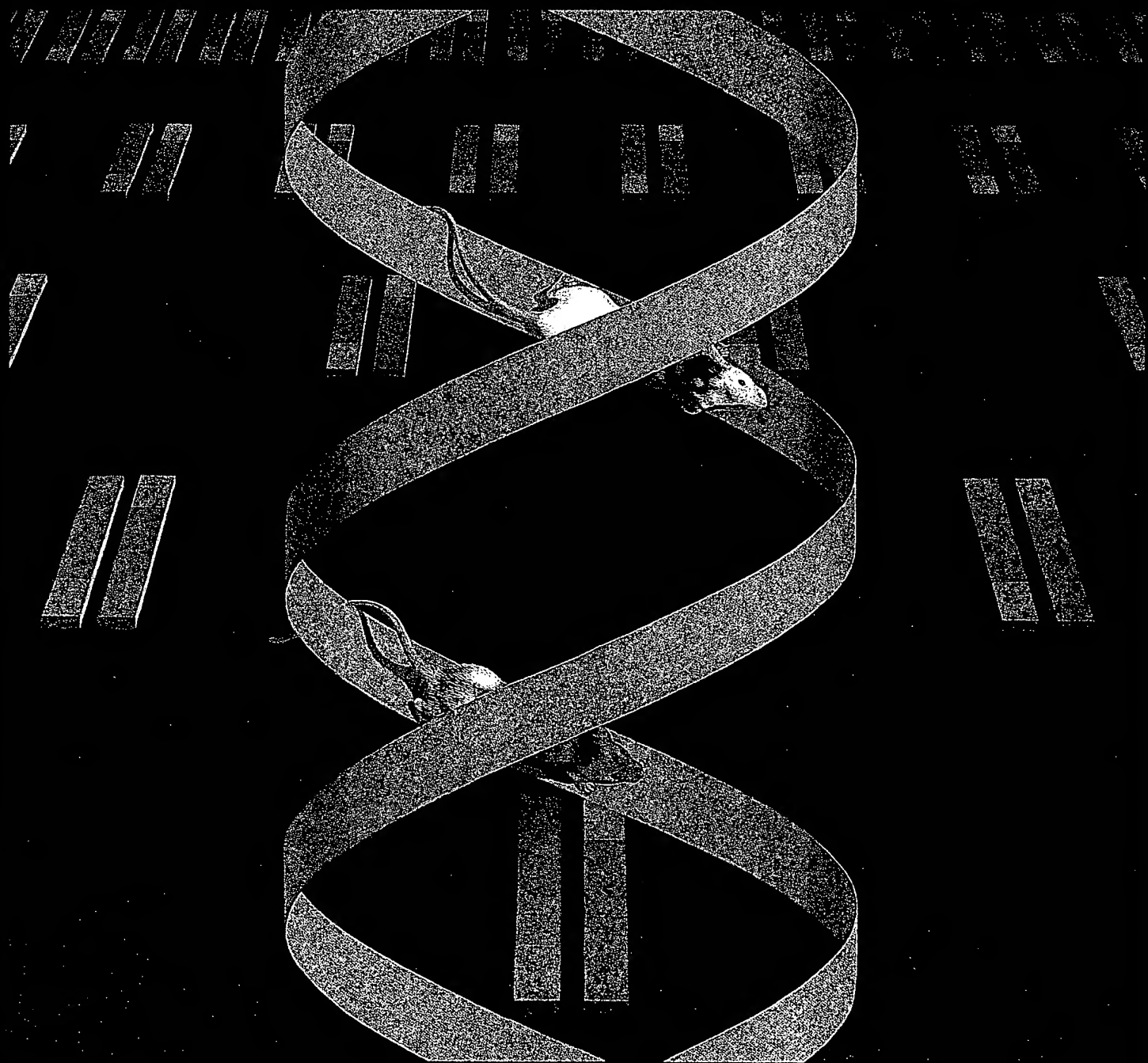
APPENDIX B

Recombinant DNA

James D. Watson

Michael Gilman

Jan Witkowski

Mark Zoller

Recombinant DNA

SECOND EDITION

James D. Watson
COLD SPRING HARBOR LABORATORY

Michael Gilman
ARIAD PHARMACEUTICALS, INC.

Jan Witkowski
BANBURY CENTER, COLD SPRING HARBOR LABORATORY

Mark Zoller
ARIAD PHARMACEUTICALS, INC.

The cover illustration, by Marvin Mattelson, symbolizes some of the elements of this book. The DNA double helix is, of course, central to the book, as it is to the cover illustration. The blocks are double-stranded DNA fragments synthesized by the polymerase chain reaction, a technique that has revolutionized the way molecular genetics experiments are done. The number of fragments doubles repeatedly, going off into the distance (see Chapter 6). The coat colors of the mice running down the helix (in the same direction but with opposite polarity!), are changing from albino to chimeric, then chimeric to agouti. These coat color changes show mice in which genetic engineering has been used to knock out a specific gene. The experiment is shown more realistically in Figure 14-9.

Printed in the United States of America

Degenerate PCR primers based
on urate oxidase amino acid
sequence

PCR

DNA fragments
made by PCR

Single-stranded cDNA
from pig liver mRNA

Select fragments of
predicted size (100 bp)

Ligate to
cloning vector

Transform
E. coli

Screen colonies by
hybridization to internal
guessmer probe

Isolate plasmid
DNA from
positives

PCR primer sequence

Insert contains urate
oxidase coding sequence

PCR primer sequence

Use insert as probe to
screen a cDNA library

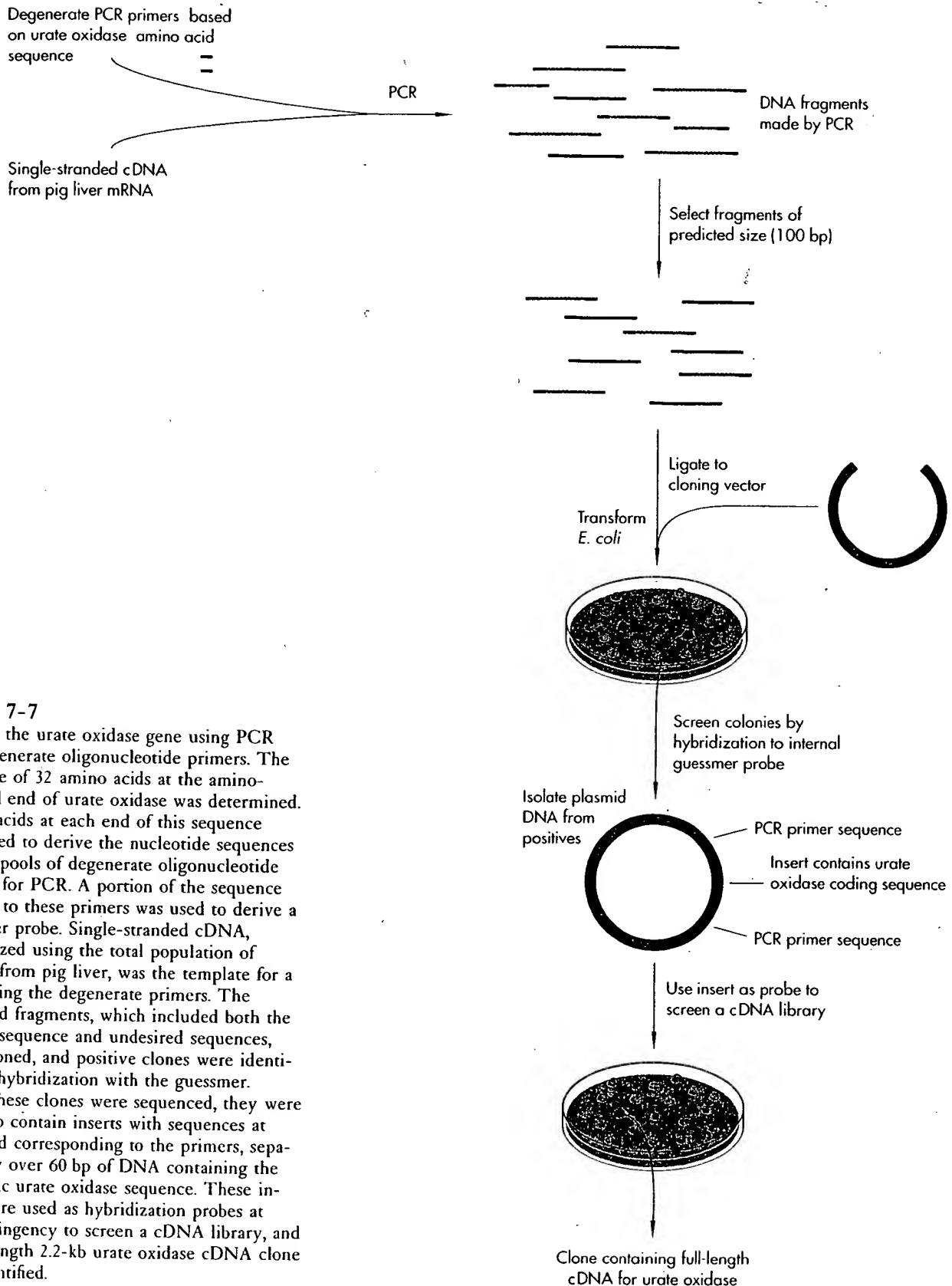Clone containing full-length
cDNA for urate oxidase

FIGURE 7-7

Cloning the urate oxidase gene using PCR
and degenerate oligonucleotide primers. The
sequence of 32 amino acids at the amino-
terminal end of urate oxidase was determined.
Amino acids at each end of this sequence
were used to derive the nucleotide sequences
for two pools of degenerate oligonucleotide
primers for PCR. A portion of the sequence
internal to these primers was used to derive a
guessmer probe. Single-stranded cDNA,
synthesized using the total population of
mRNA from pig liver, was the template for a
PCR using the degenerate primers. The
amplified fragments, which included both the
desired sequence and undesired sequences,
were cloned, and positive clones were identi-
fied by hybridization with the guessmer.
When these clones were sequenced, they were
found to contain inserts with sequences at
each end corresponding to the primers, sepa-
rated by over 60 bp of DNA containing the
authentic urate oxidase sequence. These in-
serts were used as hybridization probes at
high stringency to screen a cDNA library, and
a full-length 2.2-kb urate oxidase cDNA clone
was identified.